

AMENDMENTS TO THE SPECIFICATION:

Please replace the paragraph beginning on line 9 of page 18, with the paragraph below.

No new matter is added.

The method uses sophisticated, but fast, term recognition software to characterize documents by a list of the major, salient terms in each of the documents. When a search is performed, the terms in the returned documents can either be recognized rapidly, as described in Fig. 2, or these terms can be precomputed when the documents are indexed for retrieval and stored in the same database. Two documents are identical if they have all of the major terms identical, and very nearly identical if the number of different terms is less than some threshold value, such as 10%, of the total number of terms. A first document may be declared to be substantially identical to, or substantially similar to, a second document if some predetermined number of terms are found in both the first document and the second document. For example, if the predetermined number is about 90% of the terms or greater, the first document is declared to be substantially identical to the second document.